DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Hitchhiking and Recombination in Birds: Evidence from Mhc-linked and Unlinked Loci in Red-winged Blackbirds (Agelaius phoeniceus)

Citation
Edwards, Scott V., and Megan Dillon. 2004. Hitchhiking and recombination in birds: Evidence from Mhc-linked and unlinked loci in Red-winged Blackbirds (Agelaius phoeniceus). Genetical Research 84(3): 175-192.

Published Version
http://dx.doi.org/10.1017/S0016672304007189

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:2666813

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Genet. Res., Camb. (2004), **84**, *pp.* 175–192. *With 5 figures.* © 2004 Cambridge University Press
DOI: 10.1017/S0016672304007189 Printed in the United Kingdom

Hitchhiking and recombination in birds: evidence from *Mhc*-linked and unlinked loci in Red-winged Blackbirds (*Agelaius phoeniceus*)

SCOTT V. EDWARDS[1,2]* AND MEGAN DILLON[1]†
[1] *Department of Biology, University of Washington, Seattle, WA 98195, USA*
[2] *Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138 USA*

(*Received 20 February 2004 and in revised form 8 September 2004*)

Summary

Hitchhiking phenomena and genetic recombination have important consequences for a variety of fields for which birds are model species, yet we know virtually nothing about naturally occurring rates of recombination or the extent of linkage disequilibrium in birds. We took advantage of a previously sequenced cosmid clone from Red-winged Blackbirds (*Agelaius phoeniceus*) bearing a highly polymorphic *Mhc* class II gene, *Agph-DAB1*, to measure the extent of linkage disequilibrium across ~40 kb of genomic DNA and to determine whether non-coding nucleotide diversity was elevated as a result of physical proximity to a target of balancing selection. Application of coalescent theory predicts that the hitchhiking effect is enhanced by the larger effective population size of blackbirds compared with humans, despite the presumably higher rates of recombination in birds. We surveyed sequence polymorphism at three *Mhc*-linked loci occurring 1·5–40 kb away from *Agph-DAB1* and found that nucleotide diversity was indistinguishable from that found at three presumably unlinked, non-coding introns (β-actin intron 2, β-fibrinogen intron 7 and rhodopsin intron 2). Linkage disequilibrium as measured by Lewontin's *D*' was found only across a few hundred base pairs within any given locus, and was not detectable among any *Mhc*-linked loci. Estimated rates of the per site recombination rate ρ derived from three different analytical methods suggest that the amounts of recombination in blackbirds are up to two orders of magnitude higher than in humans, a discrepancy that cannot be explained entirely by the higher effective population size of blackbirds relative to humans. In addition, the ratio of the number of estimated recombination events per mutation frequently exceeds 1, as in *Drosophila*, again much higher than estimates in humans. Although the confidence limits of the blackbird estimates themselves span an order of magnitude, these data suggest that in blackbirds the hitchhiking effect for this region is negligible and may imply that the per site per individual recombination rate is high, resembling those of *Drosophila* more than those of humans.

1. Introduction

The increase in genome-scale data has made many previously intractable population genetic problems become more amenable to study using DNA sequence data. One such problem which was articulated decades ago (Felsenstein, 1974; Maynard-Smith & Haigh, 1974; Thomson, 1977) but which has only recently been subjected to rigorous quantitative analysis is the phenomenon of genetic hitchhiking. Hitchhiking arises when the evolutionary trajectory of selectively neutral regions in the genome is influenced by the trajectories of closely linked loci that are the actual targets of selection. Although many factors are involved, the degree of hitchhiking largely depends upon the recombination rate and the physical distance between a neutral locus and its linked selected gene, where recombination increasingly decouples the evolutionary histories of the two loci, causing the effects

* Corresponding author. Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA.
† Present address: Department of Pediatrics, Division of Immunology, Box 356320, University of Washington, Seattle, WA 98195, USA.

of selection on the target locus to be minimized at the neutral locus (Barton, 2000). Thus, estimating the extent of recombination in DNA sequences can contribute to predictions of the extent of genome-or chromosome-wide hitchhiking.

The resurgence of interest in hitchhiking is partly due to the discovery of a correlation between genomic diversity and local recombination rate in *Drosophila* and humans (Aquadro *et al*., 1994; Nachman, 2001, 2002; Payseur & Nachman, 2002*b*), a correlation which is presumably modulated by hitchhiking of neutral regions on either positively or negatively selected loci (but see Lercher & Hurst, 2002; Hellmann *et al*., 2003). Because it has the potential to amplify the effects of selection at a single site across large regions of chromosomes, hitchhiking and recombination can play a major role in influencing naturally occurring levels of genetic variation, haplotype structure and genome evolution (Posada *et al*., 2002). Indeed, another means of evaluating the extent of hitchhiking is by examining patterns of linkage disequilibrium (LD) between sites thought to be under selection and adjacent regions. Recently, discrepancies between levels of LD and estimated recombination rates have suggested that additional factors, such as population structure and selection, could also play important roles in modulating LD and haplotype structure (Pritchard & Przeworski, 2001; Stumpf & McVean, 2003). What is even less clear is how prevalent hitchhiking is in species that are not models for genetic analysis.

Few areas of the genome provide conditions more conducive to hitchhiking as a mode of evolution than the major histocompatibility complex (*Mhc*), a large multigene family encoding proteins that play an integral role in the adaptive immune response (Klein *et al*., 1993; Edwards & Hedrick, 1998; Yeager & Hughes, 1999). The high degree of polymorphism found in *Mhc* genes has been attributed to balancing selection, in the form of either heterozygote advantage or frequency-dependent selection (Garrigan & Hedrick, 2003). This selection is thought to be mediated by the rapidly evolving parasites that are bound by *Mhc* gene products, and in recent years several convincing examples of *Mhc*–parasite associations in humans and other vertebrates have been documented (reviewed in Hill, 1998; Dean *et al*., 2002). The human *MHC* (HLA) has long been a focus of studies of linkage disequilibrium, and several studies have found that the LD detected is probably maintained not so much by lack of recombination in the region but by selection maintaining distinct multilocus haplotypes (Begovich *et al*., 1992, 2001; Hollenbach *et al*., 2001). Regions of the genome surrounding *Mhc* genes have been characterized in humans as having several megabase-long tracts of DNA in linkage disequilibrium (Jeffreys *et al*., 2001; Walsh *et al*., 2003), as well as discrete hotspots of recombination (Crawford *et al*., 2004; Ptak *et al*., 2004). By contrast, analysis of non-coding regions in the vicinity of human and mouse *Mhc* genes shows that the effect of hitchhiking can be substantial, sometimes raising the level of nucleotide diversity 10-fold over the 'background' level found at neutral regions unlinked to the *Mhc* (Meager & Potts, 1997; Satta *et al*., 1999).

For a variety of structural and evolutionary reasons, the *Mhc* of birds, in contrast to human *Mhc*, may conform to an alternate evolutionary model. At less than 100 kb, the *Mhc* (B-complex) in chickens is much smaller and more gene-dense than that in mammals (Kaufman, 1999; Kaufman *et al*., 1999). This fact alone might indicate that the *Mhc* of birds would be much more subject to hitchhiking phenomena, both because of its small size and because of the high density of potential targets of selection, in the form of coding regions (McQueen *et al*., 1998; Payseur & Nachman 2002*a*). In addition, the chicken *Mhc* is located on one of the microchromosomes, which are now known to be about 3–12 Mb in length (reviewed in Burt, 2002). This characteristic could potentially increase the extent of hitchhiking for the above reasons, but could also decrease hitchhiking because of the obligatory crossing over that takes place on all avian microchromosomes every meiosis. Consistent with this idea is the finding that physical and genetic linkage data in chickens places 1 cM between on average 350 and 440 kb (Smith & Burt, 1998), whereas the average value for humans is 1 Mb per 1 cM.

In chickens, the *Mhc* appears to undergo very little recombination as evidenced by the stability of multilocus haplotypes in surveys of thousands of meioses in pedigrees (Hala *et al*., 1988). Indeed, levels of nucleotide variation at some *Mhc*-linked loci appear high, although there has been little quantification of levels of diversity or comparison with regions unlinked to the *Mhc* in chickens (Kaufman, 1995, 1999; Kaufman & Salomonsen, 1997). By contrast, analysis of sequence diversity in a chicken class I locus revealed small-scale evidence for recombination, although no attempt was made to estimate rates (Hunt *et al*., 1993). Thus existing evidence is contradictory and it is unclear whether the higher recombination rates in birds or the avian *Mhc* may ultimately render hitchhiking a weak force in moulding genetic diversity. Although surveys of haplotype stability in chicken pedigrees are a powerful method for detecting recombination, an alternative approach is the analysis of DNA sequence data in outbred populations, which has the advantage of averaging over many more thousands of meioses in the history of a species, even with relatively small sample sizes. Despite the completion of the chicken genome project, significantly

we know of no study that has systematically investigated the extent of genetic hitchhiking in chickens or any other bird, and the applicability of findings from chickens to other avian species is not known.

Among the passerine birds (Passeriformes), the *Mhc* genes of Red-winged Blackbird (*Agelaius phoeniceus*) are particularly well studied. Red-winged Blackbirds is an ecologically well studied species belonging to an evolutionarily well studied New World clade (Freeman & Zink, 1995; Beletsky, 1996). We have cloned three full-length *Mhc* class II genes from this species, only one of which, *Agph-DAB1*, exhibits the high polymorphism and evidence for selection expected of functional *Mhc* genes (Edwards *et al*., 1998). A first look at polymorphism in the intron adjacent to the *Agph-DAB1* peptide-binding region (PBR) revealed that the level of intron diversity was considerably lower than that found in the PBR, although this level was still substantial (Garrigan & Edwards, 1999). Paradoxically, this intron exhibited a signature (a significantly negative value of Tajima's *D* statistic (Tajima, 1989), of directional selection, or possibly of population growth, rather than the strongly positive Tajima's *D* expected of a locus linked to one under balancing selection. Although estimates of recombination rates in this gene were high, the absence of estimates of both neutral levels of nucleotide diversity made it difficult to determine the evolutionary forces creating the discordance in polymorphism between the two loci. Additionally, the situation upstream of the *Agph-DAB1* PBR could in principle be dramatically different from that exhibited by the downstream intron that was investigated.

The sequencing and detailed characterization of the cosmid (RWCos3) on which *Agph-DAB1* is found (Gasper *et al*., 2001) paves the way for extending the study of hitchhiking in songbirds. The present study aims to quantify the importance of hitchhiking in blackbirds by measuring diversity as well as recombination rates throughout the 45 kb region upstream of *Agph-DAB1* and by comparing these to levels of diversity and recombination found in three additional nuclear introns, presumably unlinked to the blackbird *Mhc*. Because of their large effective population sizes and lack of significant genetic structure (Ball *et al*., 1988; Garrigan & Edwards, 1999), Red-winged Blackbird is an appropriate species in which to relate evolutionary forces such as recombination specifically to LD and hitchhiking; in addition to being influenced by recombination, LD is known to be caused by bottlenecks, population structure and other demographic forces, and so studying LD in a species with little structure can help reduce the possible explanations for the patterns of disequilibrium found (Slatkin & Wiehe, 1998; Reich *et al*., 2002; Wakeley & Lessard, 2003). Finally, our data add to a small but growing data base of measurements of non-coding single-nucleotide polymorphism (SNP) diversity in birds (Congdon *et al*., 2000; Brumfield *et al*., 2003).

2. Methods

(i) *Quantitative model of hitchhiking*

We used equations 1–7 of Takahata & Satta (1998) to explore quantitative predictions for the extent of hitchhiking in a typical avian species. These equations permit a prediction of the nucleotide diversity at a neutral locus ($\pi = 2\,\mu T$, where μ is the neutral mutation rate per site per generation and T is the coalescence time in generations) separated from a target under strong symmetrical balancing selection (e.g. an *Mhc* PBR) of strength s by a rate of recombination c scaled by the sample size. The units of c are recombination events per site per generation per individual. For these analyses we relied heavily on estimates for population genetic parameters for Red-winged Blackbird (Garrigan & Edwards, 1999) and chicken (Smith & Burt, 1998). Briefly, we calculated T as:

$$T = \frac{1}{n} T_w + \left(1 - \frac{1}{n}\right) T_b, \tag{1}$$

where

$$T_w = \frac{2N_e(a + c^*n)}{(n + 2N_e a)(a + c^*)} \tag{2}$$

and

$$T_b = T_w + \frac{n-1}{2(a + c^*)} \tag{3}$$

and n is the sample size of alleles, μ is the neutral mutation rate and T_w and T_b are coalescence times within and between 'allelic lineages'. These coalescence times within and between lineages were originally parameterized for the human HLA, but are used here for simplicity. The rate of loss of allelic lineages, and hence of nucleotide diversity, is captured in the parameter a, which summarizes the relative strength of balancing selection versus genetic drift due to finite N_e, or effective population size (see Takahata & Satta (1998) equation 1). μ was centred around 10^{-9}, a value that receives some support from avian studies (Axelsson *et al*., 2004), but sensitivity to this parameter was gauged by varying it from 10^{-8} to 10^{-10} (see legend to Fig. 2). We generated a range of curves corresponding to values for N_e, from humans (10 000) and blackbirds ($\sim$50 000–100 000 based on previous mitochondrial DNA studies: Ball *et al*., 1988). The selection intensity (s) was varied such that

Table 1. *Sequence statistics for seven loci in Red-winged Blackbird*

	Mhc-linked loci					Unlinked to *Mhc*		
	Agph-DAB1 exon 2	*Agph-DAB1* intron 2	*Agph-DAB1* promoter	*Agph-DAB3*	40 kb locus	β-fibrinogen (*BFIB7*)	β actin (*BACT2*)	Rhodopsin intron 2 (*RDP2*)
No. of chromosomes	33	33	18	18	18	18	30	18
Base pairs sequenced	89	245	343	267	525	422	329	239
Base composition:								
A : C	23·1 : 30·4	16·7 : 45·0	13·1 : 35·9	22·5 : 28·1	28·7 : 26·2	35·8 : 18·8	19·0 : 23·5	25·1 : 33·9
G : T	30·7 : 15·9†	14·6 : 23·7	30·7 : 20·4	31·5 : 17·9	19·8 : 25·3	16·8 : 28·6	23·6 : 33·8	28·0 : 13·0
Ti/Tv, α	1, 0·31	1, 0·17	1, ∞	1, ∞	7·3, 0·28	2·0, 8·67	3·1, 0·36	1, ∞
Var. sites (S)	22	36	3	2	14	18	8	2
No. of haplotypes[a]	8	23	4	3	11	12	11	3
Watterson's θ	0·071	0·040	0·00254	0·0022	0·00775	0·01240	0·00614	0·00245
Nuclear diversity (π)	0·101	0·018	0·00217	0·00152	0·00998	0·01234	0·00444	0·00093
Tajima's *D*	1·9*	−1·7*	−0·40407	−0·73968	1·08303	−0·22124	−0·84182	−1·50776

Ti/Tv and α indicate the transition/transversion ratio and among-site rate heterogeneity for the locus as determined in MODEL TEST.
†Not significantly different from equal base frequencies as determined by likelihood ratio tests in MODEL TEST; $*P<0{\cdot}05$.
[a] As determined by the computer program PHASE.

the product *2Ns* for the human curves in Fig. 2 ranged between 200 and 20 000, values that encompass the range of values receiving support from recent analyses of HLA data (Satta *et al*., 1994; Slatkin & Muirhead, 2000). The same range of values for *s* (0·01–0·1) was applied to the avian curves in Fig. 2, as well as $s=0{\cdot}001$ to explore the consequences of weaker selection in birds. We applied an average physical length per centimorgan of 350 kb in chickens (Smith & Burt, 1998) to the blackbird data set. While this assumption is surely naïve, it is a good starting point.

(ii) *DNA samples*

Blackbird tissue samples were obtained from the Genetic Resources Collection of the Burke Museum, University of Washington, Seattle, with the exception of two samples from Florida from Garrigan & Edwards's (1999) study. Following a proteinase K tissue digestion, DNA was extracted using a standard phenol-TE dialysis method (Sambrook & Russell, 2001). Although sample sizes varied depending on the locus being studied (Table 1), for most loci in this study 18 chromosomes (9 individuals) were sampled. To encompass as much of the geographic range of this species as possible, we used birds from throughout North America (Washington state, Colorado, Oregon, Florida, Kentucky and Nicaragua; Appendix).

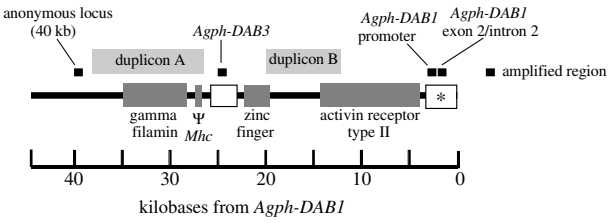

Fig. 1. Schematic diagram of cosmid RWCos3 (Gasper *et al*., 2001). Black bars above genes indicate *Mhc*-linked regions amplified.

(iii) *Genetic regions and molecular protocols*

Sequencing of the cosmid containing *Agph-DAB1* (RWCos3) revealed that the gene is located on one end of the clone and allowed us to characterize in detail the genomic region on the side of the gene opposite that of intron 2 (Fig. 1; Gasper *et al*., 2001). The sequence of this cosmid spans 45 kb and encodes an additional *Mhc* class II B gene, *Agph-DAB3*, as well as three other likely functional genes. RWCos3 was also found to have large duplicated regions paralogous to regions on a second cosmid, RWCos10 (Edwards *et al*., 2000). The new loci chosen for this study were confined to those regions known not to be duplicated or paralogous. The first region consisted of the previously published *Mhc* class II exon and intron data from the *Agph-DAB1* locus (Garrigan & Edwards, 1999; accession numbers

AF105404–AF105420 and AF133292–AF133297). The next three regions are defined as *Mhc*-linked and occur on the cosmid RWCos3 along with the *Agph-DAB1* locus (Fig. 1). These included an anonymous, apparently non-coding region 40 kb upstream of *Agph-DAB1* ('40 kb locus'; a BLAST search of this locus revealed no significant hits except the cosmid from which it came); exon 2 of *Agph-DAB3* ('*DAB3*'), which was also surveyed for polymorphism in Gasper *et al.* (2001), but for a different set of individuals; and the promoter regionof *Agph-DAB1* ('*DAB1* promoter'). The primers 3RW-1035F (5′-TACTGGTAGCAACAGCCATT-3′) and 3RW-1934R (5′-AGCAGCTAATAGCCATGCTT-3′) amplified the 40 kb locus; primers in Gasper *et al.* (2001) were used to amplify *DAB3*; and primers DAB1-RR-F (5′-GCTGGATGATCCCCAGTGAC-3′) and DAB1-RR-R(5′-TCTCATTGGCTGCAACGCCTC-3′) amplified the *DAB1* promoter. Additionally, three non-coding loci, presumably unlinked to the blackbird *Mhc*, were amplified: intron 2 of rhodopsin ('*RDP2*'), intron 7 of the β-fibrinogen gene ('*BFB7*') and intron 2 of the β-actin gene ('*BACT2*'), all of which have been used previously for avian evolutionary studies and primes for which are available (Prychitko & Moore, 1997; Waltari & Edwards, 2002).

All regions were directly sequenced on an ABI 377 automated sequencer with the above amplification primers with the exception of the 40 kb locus, which was directly sequenced with the primers 3RW-1188F (5′-TCTCCAACAGAGTGCGATAA-3′) and 3RW-1840R (5′-ACCTCTGTGGTCCCTAGGTT-3′). In order to conduct HKA tests of neutrality, which require an outgroup species (Hudson *et al.*, 1987), we attempted to amplify these loci from DNA from Brown Cowbird (*Molothrus ater*) and Bullock's Oriole (*Icterus bullocki*). Because only the *RDP2* and *BFIB7* loci amplified and sequenced reliably from these species, and because the HKA tests conducted did not reject neutrality, we do not report data from these tests further. All new sequences (those unlinked to the Mhc) have been submitted to the GenBank data base under accession numbers AY714388 (*RDP2*), AY714390 (*BFIB7*), AY714391 (*BACT2*). SNPs in these and the Mhc-linked loci have been submitted to the NIH dbSNP data base under SNP accession numbers 28514110–28514117 (*BACT2*) and 28514123–28514124 (*RDP2*), 28514125–28514127 (*DAB1* promoter), 28514128–28514129 (*DAB3*), 28514130–28514147 (*BFIB7*), 28514148–28514161 (40 kb).

(iv) *Measures of polymorphism and neutrality*

As an aid to manual sequence inspection in the search for polymorphic SNPs, the raw chromatograms were analysed using Polyphred (Nickerson *et al.*, 1997; Rieder *et al.*, 1998), which uses details of chromatogram peak height, morphology and quality to identify heterozygous sites in diploid PCR products (see Brumfield *et al.*, 2003). Genotypes of heterozygous SNPs were then resolved into haplotypes using PHASE v. 2.0, a program that yields Bayesian estimates of haplotypes and their frequencies from genotypic data under the assumption of random mating (Li & Stephens, 2003; Stephens & Donnelly, 2003). Phases of the exon 2/intron 2 data of Garrigan & Edwards (1999) had already been resolved via cloning of PCR products; in the case of the three additional *Mhc*-linked loci analysed here, we incorporated the phase information from one individual (individual 7, Appendix) provided by the original cosmid sequence into our PHASE analysis. Haplotypes resolved using PHASE generally had high probabilities associated with heterozygous sites and were robust to assumptions of recombination or lack thereof (Appendix). The program DNAsp (Rozas & Rozas, 1997) was used to calculate basic sequence statistics, such as nucleotide diversity (π) and Tajima's *D*. Although no phylogenetic analyses were done, we nonetheless used PAUP*4.0b10 (Swofford, 1999) and the program MODELTEST (Posada & Crandall, 2001) to collect basic evolutionary information on these predominantly non-coding loci, such as base composition, transition – transversion ratio and rate heterogeneity among sites (Table 1).

(v) *Linkage disequilibrium and recombination*

Haplotypes generated using PHASE were used to calculate *D'*, a common measure of linkage disequilibrium (Lewontin, 1995), between polymorphic sites of the *Mhc* linked loci using the program LD plot, provided by S. Schaeffer (see Schaeffer *et al.*, 2001). Because the set of individuals used by Garrigan & Edwards (1999) and in this study were only partially overlapping, we only estimated LD for sites within each of these two data sets. Because of our small sample size for some loci, we examined statistical significance with and without Bonferroni correction (Rice, 1989).

As recommended (Posada & Crandall, 2001), we used multiple computer programs to estimate the population recombination parameter, $\rho = 4Nc$ from the collected sequence data (for recent reviews see Posada, 2002; Rokas *et al.*, 2003). The program Ldhat (Hudson, 2001; McVean *et al.*, 2002) is a maximum likelihood method that calculates the coalescent likelihood of every pair of polymorphic sites, and generates a composite likelihood function of ρ per locus across all sites. ρ per site was calculated by dividing the Ldhat ML estimate by the number of sites in the locus. We also used PHASE to estimate ρ per site while simultaneously estimating haplotypes

(Stephens *et al.*, 2001*b*; Li & Stephens, 2003; Stephens & Donnelly, 2003). PHASE generates a posterior probability distribution of ρ and can be checked for convergence. When estimating ρ we ran PHASE with varying random number seeds and starting points for ρ, usually using the commonly cited value from humans of 0·0004 per site (equivalent to 1 recombination event per million base pairs per generation). Each run consisted of a large number of iterations, usually 10 000. Both PHASE and Ldhat analyse only single sets of linked sites at a time. Finally, we used the Markov Chain Monte Carlo method RECOMBINE (Kuhner *et al.*, 2000), as implemented in the program LAMARC (http://evolution.genetics.washington.edu/lamarc.html), to estimate ρ for individual loci as well as to generate an overall value of ρ across all loci analysed, as well as associated values of $\theta = 4\,N\mu$. Again, multiple runs and increases in Markov chain length and number were used to ensure convergence of parameter values. Although all of these programs assume neutrality in their estimation of ρ from sequence data, we nonetheless analysed some loci clearly violating neutrality (e.g. *Agph-DAB1*) for comparison with loci determined to be evolving neutrally. Two types of multilocus analyses were done. First, the three *Mhc*-linked loci spanning ~40 kb were analysed simultaneously using PHASE and Ldhat; in the latter procedure, the average value of θ across the three loci was applied to the entire ~40 kb region. Second, a multilocus average value for ρ was estimated using LAMARC. Here, as many loci as possible were included, regardless of level of polymorphism, so as to maximize precision of the global estimate.

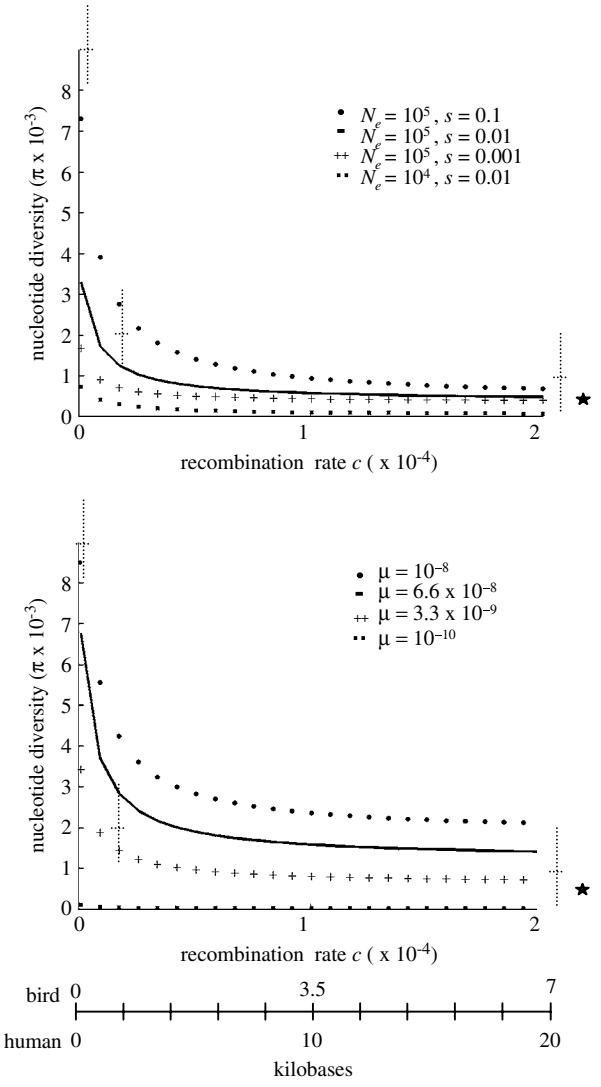

Fig. 2. Graphical analysis of nucleotide diversity versus recombination rate according to equations in Takahata *et al.* (1998). Parameter values for each curve appear at upper right. In both panels the line indicated by squares corresponds to parameters found in humans, whereas the three additional lines correspond to possible values of parameters for blackbirds. Scale showing expected physical distance for putative human (1 cM/Mb) and chicken (1 cM/350 kb; Smith & Burt 1998) appears at the bottom. The observed values of π for three *Mhc*-linked loci in this study (left to right: *Agph-DAB1* exon, *Agph-DAB1* promoter, *Agph-DAB3*; Table 1) and 2 standard errors are indicated in dashed lines at the appropriate location on the *x*-axis for comparison with theoretical expectations. A star to the right of each panel indicates the expected value of $4\,N\mu$ when $N = 10^5$ and $\mu = 10^{-9}$.

3. Results

(i) *Theoretical expectations*

Application of the hitchhiking model of Takahata & Satta (1998) and recombination parameters from chickens (Smith & Burt, 1998) suggest that the hitchhiking effect should extend along the chromosome to similar degrees in birds and humans. In our exploration of parameter space, we varied only N_e and s. Like Takahata & Satta (1998), we found that the product Ns was largely responsible for setting the leftmost 'height' of the curve and the level of neutral diversity close to the target of selection (on the left-hand end of Fig. 2), whereas N_e alone influenced the level of diversity far from the target of selection (on the right-hand end of Fig. 2). The values corresponding to human populations (lowest curve) fell to neutral levels within about 6 kb, whereas, depending on the values of N_e and s assumed for the avian curves, nucleotide diversity could remain higher than the neutral expectation at least this far from the target of selection. N_e seemed to influence the rate of decline to neutral levels more than did the recombination rate.

(ii) *Polymorphism and haplotyping*

Of the 1135 nucleotides resequenced across the three *Mhc*-linked loci (*DAB1* promoter, *DAB3*, 40 kb

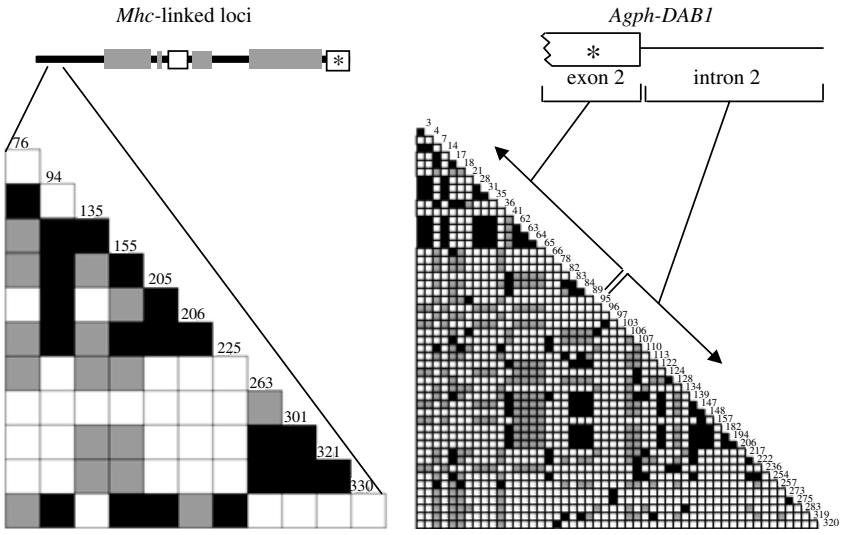

Fig. 3. Patterns of linkage disequilibrium at variable sites for *Mhc*-linked loci (left) and *Agph-DAB1* (right). Numbers above the diagonal indicate each variable site in the locus using the numbering system in the Appendix. Black boxes indicate pairs of sites exhibiting significant value of Lewontin's (1995) *D*' without Bonferroni correction. Grey boxes indicate non-significant associations of sites for which calculation of *D*' is possible. Schematics of gene regions are shown above each matrix.

locus), 19 sites were polymorphic, yielding an overall θ_w (Watterson's θ) of 0·006, a value that is substantially higher than that found in humans but similar to levels in other birds (Wang *et al*., 2003; see Table 1). Across 990 sites resequenced for the three introns unlinked to the *Mhc* (*BFIB7*, *BACT2*, *RDP2*), there were a total of 28 polymorphic sites, yielding an overall θ_w of 0·007, which is indistinguishable from the average value for the *Mhc*-linked loci. θ_w for both *Mhc*-linked and presumably unlinked loci were substantially lower than those for the *Agph-DAB1* exon 2 and intron 2 (0·101, 0·018, respectively). Nucleotide diversity (π) was similar to θ_w for all loci, except for the *Agph-DAB1* intron, which showed a smaller π value given θ_w and therefore rejected neutrality (Garrigan & Edwards, 1999). Overall *P* values were consistent with theoretical expectations resulting from parameter combinations of high N_e and/or high *s* (Fig. 2). For the two protein-coding loci synonymous diversity (π_S) was similar to non-synonymous diversity (π_N), although both these values for *Agph-DAB3* were small (DAB3: π_S, 0·00172; π_N, 0·00146; *Agph-DAB1*. None of the other six loci rejected the hypothesis of neutrality using Tajima's *D* test and only one locus (40 kb) was positive for Tajima's *D* (Table 1). The *D* value for RDP2, although non-significant, was substantial and may warrant further investigation. The interspecific divergence for these loci largely predicted the within-blackbird level of polymorphism, and none was able to reject neutrality by the HKA test.

The genotypic variation was resolved into haplotypes using PHASE, with and without the assumption of recombination. The variation in the three *Mhc*-linked loci was resolved simultaneously into 17 distinct haplotypes out of a total 18 possible, indicating high haplotypic diversity (Appendix). Although the most probable multilocus haplotypes were the same whether or not information from the two known haplotypes was included, probabilities of individual sites within haplotypes generally increased due to inclusion of these known haplotypes (not shown). The β-fibrinogen intron (*BFIB7*) also had a high haplotypic diversity, with 13 out of a possible 18 haplotypes recovered, each of which was found in only 1 or 2 copies. For the two additional loci unlinked to *Mhc* genes, there was a single common inferred haplotype and a smaller number of total haplotypes: *BACT2* had 9 inferred haplotypes, the most common of which was inferred to have occurred 12 times, and *RDP2* had 3 haplotypes, with the most common found 16 times.

(iii) *Linkage disequilibrium*

Analysis of the *Agph-DAB1* exon/intron data set of Garrigan & Edwards (1999) suggested high levels of LD within the exon, but fewer sites in the intron in LD with other sites (Fig. 3, left). Of the 50 sites exhibiting some level of LD with other sites in the region ($P<0·05$), 20 (40%) were in the exon, even though only 89 of the 337 total sites (26%) were exonic. High levels of LD in the exon were probably generated by a gene conversion event, discussed by Garrigan & Edwards (1999), that caused a suite of sites (3, 4, 7, 28, 31, 35, 62–64) to covary highly non-randomly. There

Table 2. *Estimates of recombination rate* (ρ) *in Mhc-linked and unlinked loci in Red-winged Blackbird*

	Mhc-linked			Unlinked	
	Agph-DAB1 exon 2/intron 2	40 kb locus	Multilocus[a]	β-fibrinogen intron 2 (*BFIB7*)	β-actin intron 2 (*BACT2*)
No. of sites spanned	337	525	40 900	422	329
Ldhat (ML score)	−38 002·66	−2310·03	−4885·37	−227·76	−91·34
ρ (per locus)	15·152	33·333	2789·58	94·95	36·71
ρ (per site)	0·045	0·0635	0·0682	0·225	0·112
CI (per site)[c]	0·0412–0·056	0·036–0·104	∼0·049–0·125	0·096–∞	0·0154–0·311
ρ/θ	0·96	8·19	16·39	18·14	14·01
PHASE					
ρ (per site)	0·0059–0·0088[b]	0·0767–0·0512[b]	0·0488–0·0586[b]	0·0436–0·0859[b]	0·2310–6·283[b]
95% CI	0·0016–0·0241	0·0092–0·2612	0·0082–0·2137	0·0041–90·58	$7{\cdot}22\times10^{-16}$–09·8
ρ/θ	0·127	6·604	11·726	3·517	37·626
LAMARC (ML)					
θ	n.d.	0·008–0·022[b]	0·009–0·020[b]	0·016–0·030[b]	0·006–0·012[b]
θ 95% CI	n.d.	0·007–0·021	0·008–0·011	0·001–0·029	0·004–0·011
ρ/θ	n.d.	0·066–2·372	0·353–2·496	0·125–1·735	0·123–5·927
ρ/θ 95% CI	n.d.	0·133–0·189	1·067–1·779	0·761–2·16	2·477–7·954
ρ (per site)	n.d.	0·0005–0·0524	0·003–0·049	0·002–0·051	0·0007–·0736

n.d., not done.
[a] For the Ldhat and PHASE analyses, the 'Multilocus' category refers to analysis of the *Mhc*-linked loci (40 kb, *DAB1*-promoter, *DAB3*). For the LAMARC analyses, 'Multilocus' refers to analysis of all six neutral loci (all except *Agph-DAB1* exon and intron), some of which are unlinked to *Agph-DAB1*. For the PHASE and LAMARC analyses, the 95% CIs are those recovered from a single representative run.
[b] Range of values indicates the range of ML estimates across multiple runs.
[c] CIs from Ldhat based on ML ± 2 likelihood units.

was nonetheless a cluster of sites in the intron (139, 147, 148, 182, 194) that showed substantial LD with each other and with several exonic sites (83, 84, 89) without Bonferroni correction, suggesting that, despite an inference of approximately 8 recombination events in the region by Garrigan & Edwards (1999), LD might extend throughout the region. Still, with Bonferroni correction, the region experiencing significant LD extends only to site 65 in the exon, and does not include any intronic sites.

LD was measured across the three additional loci spanning ∼40 kb of the blackbird cosmid. Significant levels of LD, however, were found primarily within the first locus, 40 kb, extending in patches throughout all 14 variable sites (Fig. 3, right) spanning 419 bp ($P<0{\cdot}05$), or across the first 5 variable sites, spanning only 96 bp with Bonferroni correction. LD ($P<0{\cdot}05$) was suggested between four sites in 40 kb (155, 263, 321, 330) and site 40868 in the *DAB1*-promoter, but these associations vanished after Bonferroni correction.

(iv) *Estimates of recombination rate* (ρ)

The program Ldhat was used to analyse five data sets. The *RDP2* and *DAB1*-promoter loci were not analysed individually because of the low level of polymorphism (2 and 3 segregating sites, respectively). Maximum likelihood estimates of ρ per site varied from a low of 0·045 for *Agph-DAB1* to 0·225 for the *BFIB7* locus (Table 2). However, the likelihood curve for *BFIB7* was very flat, indicating little confidence in the upper bound of the estimate (Fig. 4*d*). By contrast, the likelihood curve for *Agph-DAB1* was much tighter (Fig. 4*a*), with most of the signal coming from variation in the exon rather than from the intron. The number of segregating sites within a locus only partly explained the variation in likelihood curves, since *BACT2*, which had the fewest variable sites, had a detectably sharper curve than did *BFIB7* (not shown). A multilocus analysis of the three *Mhc*-linked loci yielded an estimate of $\rho=0{\cdot}0682$ that was consistent with those for most of the loci when analysed individually, although still exhibiting a very flat likelihood curve (Fig. 4*c*). Given these ML estimates of ρ, estimates of ρ/θ_w varied from less than 1 for *Agph-DAB1* to greater than 18 for the *BFIB7* locus (Table 2).

The program PHASE yielded estimates of ρ that were largely consistent with those for Ldhat, except that the relative rates of recombination for *BFIB7* and *BACT2* were reversed. The estimates of ρ for the 40 kb locus and for the multilocus data set spanning the cosmid were very similar to those of Ldhat

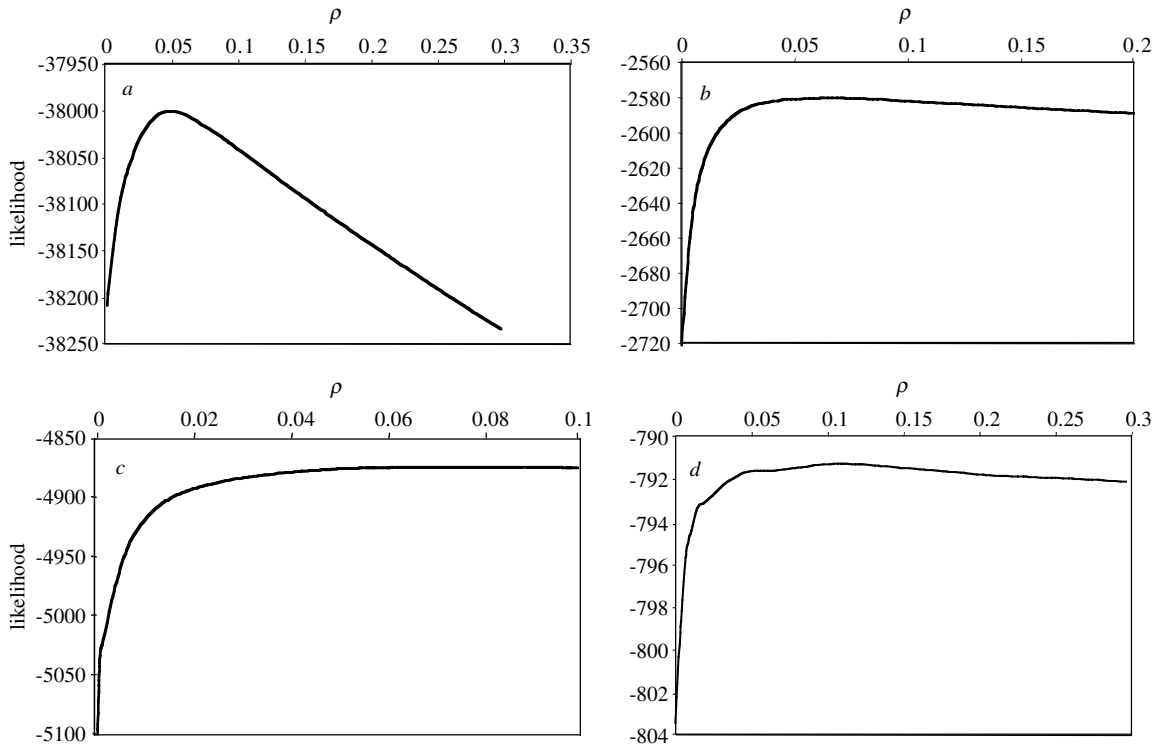

Fig. 4. Likelihood curves for estimation of ρ for four data sets produced by Ldhat. (*a*) *Agph-DAB1*; (*b*) 40 kb locus; (*c*) *Mhc*-linked loci (40 kb, *DAB3*, *DAB1* promoter); (*d*) β-actin (*BACT2*). The shape of the curve for β-fibrinogen (*BFIB7*) was very similar to that in (*d*).

(Table 2); as with Ldhat, the estimate for *Agph-DAB1* was the lowest among the loci (Fig. 4*a*, *b*), although with PHASE the estimate for this locus was an order of magnitude lower than for Ldhat. Whereas the smoothness of the likelihood surfaces for *Agph-DAB1* and the 40 kb locus were reasonable (Fig. 5*a–d*), the surface for *BACT2* was less smooth (Fig. 5*g*, *h*). The surface for *BFIB7* was distinctly bimodal, with several cycles getting caught in optima of very high values for ρ, and others converging on exceedingly low values (Fig. 5*e*, *f*). This probably indicates conflicting signals in the data, substantial homoplasies, or perhaps recombination rates that are very high and difficult to estimate. Even so, the lower bound of the 95% confidence limits for most loci excluded a commonly accepted average rate of recombination in humans of 0·0004 per base pair. As with Ldhat, all the estimates of ρ/θ_w except that for *Agph-DAB1* exceeded 1 (Table 2).

The recombination module within the LAMARC package returns estimated values of θ (θ_L) and c/μ for each locus, as well as an average estimate across all loci under the default assumption that c for all loci is equal. For this reason the loci with few segregating sites such as *RDP2* and *DAB3*, for which estimating individual recombination rates would be very difficult, were included, since data from additional loci can only improve the global estimate. The results for analyses in which haplotypes were estimated simultaneously were similar to those in which haplotypes were assumed to be known. Estimates of θ_L using LAMARC were slightly higher than Watterson's θ for most loci (Table 2). Estimates of c/μ were substantially lower than those calculated for values from Ldhat or PHASE, ranging from $\sim$0·1 to 2·5 for most loci, but as high as 5·9 for *BACT2* in one run. Values of c/μ for *RDP2*, *DAB1* promoter and *DAB3* fell outside this range ($<$0·000006; data not shown), but, as expected, there were very few segregating sites to estimate recombination for these loci singly. Estimates of ρ that were back-calculated from these values were generally lower than values from the other two methods and ranged from 0·0005 to 0·07 for those loci with reasonable information content (Table 2). The lower bound of these estimates is broadly consistent with recent estimates from humans.

We used the range of loci analysed using LAMARC to determine whether there was a correlation between θ_L and ρ. We found a strong correlation using values generated by LAMARC but a much weaker pattern using estimates from other methods (not shown).

4. Discussion

The rate and amount of recombination and extent of hitchhiking both have important consequences for

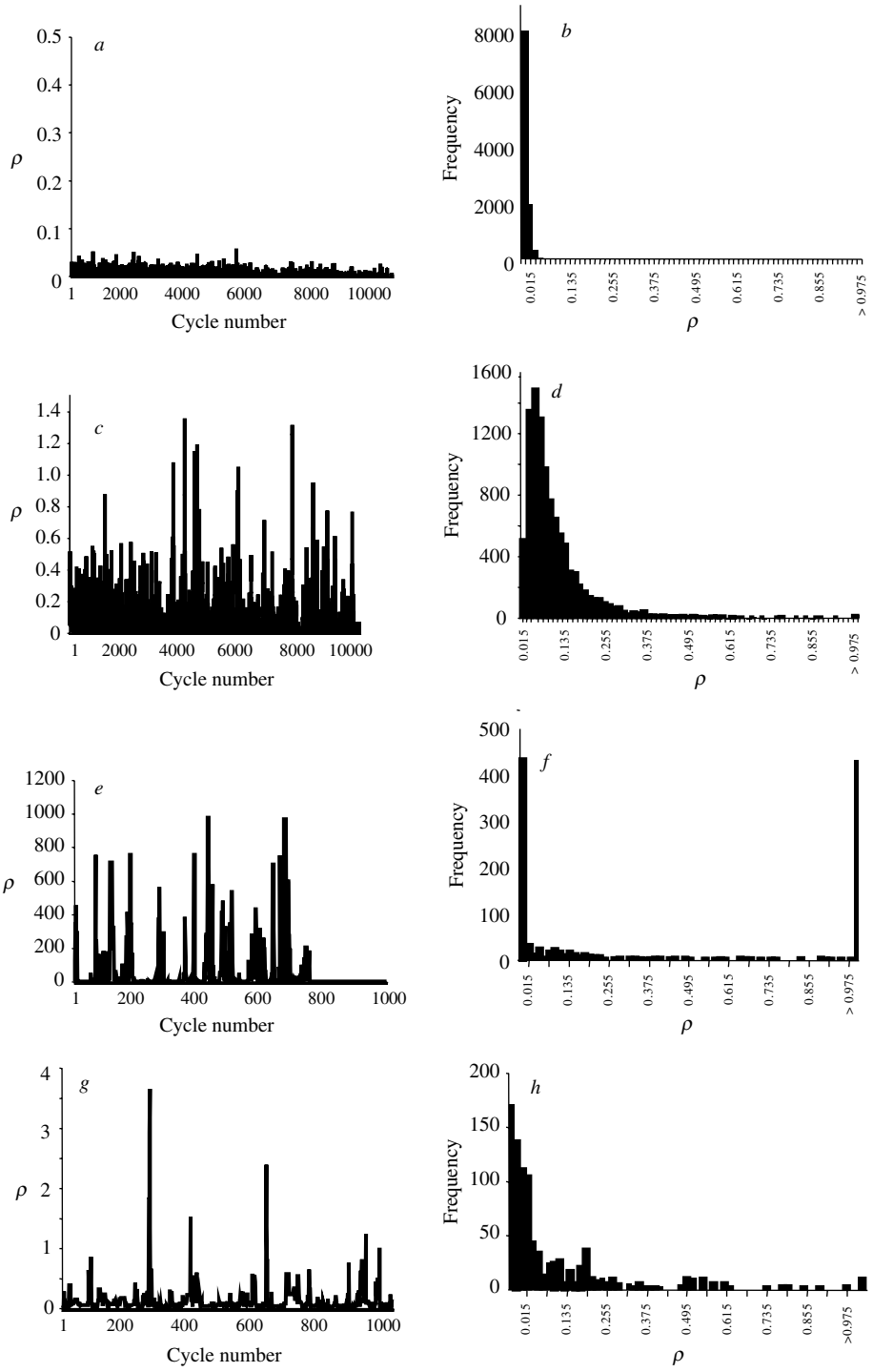

Fig. 5. Bayesian analysis of recombination using PHASE. Graphs in the left-hand column show traces of ρ values throughout the cycles of analysis. Graphs in the right-hand column show the posterior probability distribution of estimates of ρ for each data set. (*a*), (*b*) *Agph-DAB1*; (*c*), (*d*) 40 kb locus; (*e*), (*f*) *BFIB7*; (*g*), (*h*) *BACT2*.

a variety of fields (Pritchard & Przeworski, 2001), including those in which birds are important models, yet virtually nothing is known of naturally occurring levels of recombination in any bird other than chicken. For example, recombination can cause difficulties when estimating nuclear gene genealogies (Hare, 2001), but can allow non-coding portions of the genome – of increasing importance to avian population geneticists – to escape the effects of hitchhiking. In principle recombination rates also have important consequences for animal behaviour: for example, *Mhc* genes have been suggested to serve as kin-recognition or mate choice loci in outbred populations (Manning *et al.*, 1992; Jacob *et al.*, 2002; Zelano & Edwards

2002), yet the number of independently segregating *Mhc* loci will dictate the level of matching complexity achievable with a tightly linked set of kin-recognition loci (Grafen, 1990). In this paper we have explored the extent of genetic hitchhiking and used several methods to estimate rates of recombination within and between loci that are linked and presumably unlinked to an *Mhc* gene under strong balancing selection. We find that estimated amounts of recombination are high and generally exclude a commonly cited level of the number of crossovers per nucleotide per generation of 0·0004 in humans. Other estimates in humans would place $\rho = 0{\cdot}0004$ at the upper bound of recombination rates (Stumpf & McVean, 2003; McVean *et al.*, 2004), thereby accentuating the difference between blackbird and human results even more. These results are consistent with high average recombination rates in chickens but are at odds with observed levels of recombination in the *Mhc* (B-complex) in chickens.

(i) *Extent of hitchhiking in the vicinity of* Agph-DAB1

Garrigan & Edwards (1999) found slightly elevated levels of variation in the intron immediately adjacent to the PBR of *Agph-DAB1*, and this study revealed levels consistent with neutrality in the *Agph-DAB1* promoter a mere $\sim$600 bp from the PBR. Thus the levels of neutral levels of diversity predicted by analytical theory remain high much farther from the target of selection than is observed in our blackbird data. Satta & Takahata (1999) found a similar discrepancy between expected and observed extent of hitchhiking in the human HLA. The analytical theory predicted that for avian population parameters, levels of diversity should decline to neutral levels within about $\sim$7 kb away from the target of selection. Although the theory necessarily corresponds to an idealized population – in particular recombination rates are assumed to be constant thought the hypothetical *Mhc* region – we were surprised at the similarity in predictions between birds and humans (Fig. 2). It appears that the effective population size is very important in modulating the hitchhiking effect: if we assume a 10-fold higher N_e in blackbirds, even with the same selection coefficient, theory predicts that the asymptote for neutral variation is achieved at least as far from the target of selection as in humans, despite a 3-fold higher recombination rate assumed from the chicken data (Smith & Burt, 1998). Satta *et al.* (1999) collected data indicating that silent diversity can extend in excess of 100 kb from the nearest target of selection in the human HLA region, and Grimsley *et al.* (1998) found enhanced variation at an HLA pseudogene residing $\sim$200 kb from HLA-A, a target of strong balancing selection. It is likely that the close linkage of many HLA genes – some occur within 40 kb of one another – could be responsible for this maintenance of silent diversity so far from the presumed targets.

(ii) *Rates of recombination at blackbird loci*

Our sampling scheme for sequencing in this study made estimation of recombination rates challenging. Recombination is best detected across long DNA sequences that will record a number of crossover events detectable by the presence of all pairs of two-locus combinations (Posada & Crandall, 2001; McVean *et al.*, 2002). Thus, the fact that our loci were at most a few hundred base pairs long, and in some cases encompassed coding regions or regions possibly under strong stabilizing selection (*Agph-DAB1* promoter, *Agph-DAB3*), resulted in low polymorphism, making estimation of recombination difficult. On the other hand the values of θ, particularly for the non-coding loci, were generally high compared with human population studies. For nearly any method, estimating very high or very low recombination rates is difficult, because the data necessary to discriminate between multiple extreme values of ρ are usually lacking. Thus several of the loci exhibited very flat likelihood curves when analysed using Ldhat. Nonetheless, such analyses could reject very low values for ρ in regions of the space where the likelihood curve rose steeply (Fig. 4).

Our cosmid sequence and multilocus analyses in principle permitted estimation of ρ across tens of kilobases. Estimates of ρ from such analyses were surprisingly similar to those within individual loci, and both PHASE and Ldhat were able to accommodate such data easily. A recent study of recombination rates in humans (Clark *et al.*, 2003) employed similar sampling scheme with a small number of SNPs (4–10) sampled across $\sim$50 kb regions of the genome. While the human study benefits from complete knowledge of all non-polymorphic sites in such regions, most available recombination estimation packages (except LAMARC) do not take information from such invariant sites, such as base composition, into account when estimating ρ. Additionally, our sampling scheme for the *Mhc*-linked loci, which involved complete resequencing across selected loci within a larger region, avoided some of the problems associated with ascertainment bias that require special statistical procedures to ameliorate (Wakeley *et al.*, 2001; Nielsen & Signorovitch, 2003).

The seven loci we surveyed exhibited a range of recombination rates. Surprisingly, for those loci with moderate or high levels of θ, the lowest estimated rates came from the *Agph-DAB1 Mhc* locus. The frequency of recombination and/or gene conversion in mammalian *Mhc* class II loci, though controversial, is generally thought to be high (Gyllensten *et al.*, 1990; Begovich *et al.*, 1992; Bergström *et al.*, 1998;

Hogstrand & Bohme, 1999*a*, *b*), although some analyses suggest otherwise (Nei & Hughes, 1992; Gu & Nei, 1999). A recent survey of a *Mhc* class II locus in *Peromyscus* estimated recombination rates substantially higher than θ (Richman *et al*., 2003*b*). Ohta (2000) analysed data from the blackbird *Agph-DAB1* locus as well as several mammalian *Mhc* class II genes and found that the standard identity excess, a statistic that quantifies the 'patchwork' pattern resulting from frequent gene conversion and with properties similar to Tajima's *D*, was demonstrably higher for mammalian than for avian PBR exons. Visual inspection of aligned avian class II sequences confirms that the bold patchwork pattern found in, for example, mouse and *Peromyscus* class II genes (She *et al*., 1990; Richman *et al*., 2003*a*), is not as evident at avian class II loci (Edwards *et al*., 1998). Richman *et al*. (2003*b*) recently suggested that such motifs persist in rodent class II loci in the face of high recombination because of selection for combinations of motifs. It may be that recombination and gene conversion rates are in fact lower at avian than in mammalian class II loci, resulting in less of a patchwork pattern.

Throughout our study we have focused on measuring recombination in terms of the composite parameter $\rho = 4Nc$, which necessarily confounds N_e and c. Thus our finding of high values of ρ compared with humans could be consistent solely with a larger effective population size in blackbirds versus other species. The data for blackbirds do indeed suggest that N_e is higher than in humans: Ball *et al*. (1988) estimated N_e from mtDNA at about 50 000 and the values for θ_w for neutral loci in this study (0·002–0·012) imply an N_e of $\sim$400 000 if we assume an autosomal non-coding mutation rate of $3{\cdot}6 \times 10^{-9}$ as in gamebirds (Axelsson *et al*., 2004). However, even these values for N_e in blackbirds cannot entirely explain the difference in ρ between blackbirds and humans. We suspect that both N_e and c are contributing to the higher ρ in blackbirds.

(iii) *Selection and linkage disequilibrium*

Why is the estimate of ρ at *Agph-DAB1* so low relative to the neutral loci in this study? The programs used to estimate ρ all assume selective neutrality. Although the large amount of polymorphism available in *Agph-DAB1* facilitates estimation of ρ, the mode of selection on *Mhc* loci could yield spurious results with respect to estimating ρ. Epistatic interactions undoubtedly play a role in modulating *Mhc* allele frequencies, both at the level of linked genes as well as among sets of amino acids linked within the PBR (Hedrick & Thomson, 1983; Hedrick *et al*., 1991; Hunt *et al*., 1993; Gaudieri *et al*., 1999, 2000; Jacob *et al*., 2000; Jeffreys *et al*., 2001; Walsh *et al*., 2003). In particular, if epistatic interactions among PBR codons are strong, recombination rates could still be high but would not be evident in the standing patterns of diversity because haplotype blocks would be preserved (Ohta, 1999). Assuming the low estimate of ρ at *Agph-DAB1* is real, however, this argument is contradicted by the fact that the stronger block pattern of variation found in mammals, presumably driven by epistasis, is observed despite higher estimated rates of recombination. It is conceivable that stronger selection for polymorphism could be occurring in mammals; this would actually increase the amount of recombination, if recombinant alleles are more likely than *de novo* mutations to result in new allelic specificities. This argument has the virtue of being consistent with the observation of more recombination at MHC loci in mammals, and suggests the need to obtain precise estimates of selection intensities at avian loci in the future.

Slatkin (2000) has recently modelled the joint effects of selection, mutation and recombination, with the goal of describing the evolutionary forces acting on two-locus haplotypes. He points out that the standard descriptions of linkage disequilibrium may not be adequate to capture fully the dynamics acting on such systems. Because most methods of estimating recombination rely in some form on the frequencies of gametic phase haplotypes, it is likely that new approaches will be required to estimate ρ when selection cannot be ignored. It may be possible in the near future to estimate jointly selection intensities, recombination rates and population sizes from sequence data (J. Felsenstein, pers. comm.), but it is unclear how well simple selection models will approximate the forces acting on real data. In addition, demographic forces, such as population structure, can strongly influence the pattern of linkage disequilibrium, and hence, estimation of recombination rates (Navarro & Barton, 2002; Wakeley & Lessard, 2003). Population structure and the different genetic backgrounds it produces generally increase the effects of hitchhiking, increasing neutral variability in the vicinity of selected loci, although the precise consequences depend on a number of factors.

Finally, it is also unclear how much data will be necessary to estimate all these parameters jointly with precision from sequence data. For outbred species such as birds that do not yet have or are not amenable to having genetic maps, such estimation will remain a challenge, because several basic parameters available for model species, such as effective population sizes and local recombination rates, are becoming reasonably well known and do not require independent estimation with each study. Nonetheless, the sequence data we have collected in blackbirds strongly indicate that recombination rates in birds

may be substantially higher than in humans, and that the hitchhiking effect, at least for this particular genomic region in blackbirds, may be negligible. Recent estimates of recombination rates at neutral loci in *Drosophila* (Wakeley & Hey, 1997; Andolfatto & Przeworski, 2000; Kliman *et al*., 2000; Hey & Kliman, 2002; Machado *et al*., 2002) are very high and suggest that ρ/θ frequently exceed 1. Thus blackbird parameters seem to resemble those of *Drosophila* and *Peromyscus* more than those of humans. What is needed now is more nuclear DNA sequence data from loci known to be linked, in both neutral and selective contexts. Genomic approaches that allow large amounts of sequence data to be collected from multiple nuclear loci will undoubtedly play an important role in this endeavour (Stephens *et al*., 2001*a*; Clark *et al*., 2003). Although issues of recombination and hitchhiking will remain a challenge for clades such as birds, the consequences of these forces for topics for which birds are well suited will not diminish.

We are very grateful to R. Tinghitella for sequencing the *BACT2* locus; to N. Rotzel for submitting the SNP and sequence data to the data bases; and to T. Hsieh and M. Dillon for assistance with Fig. 2. We thank B. Charlesworth, P. Awadalla, M. Nachman, B. Payseur, J. Felsenstein, D. Garrigan, S. Schaeffer and A. DiRienzo for helpful discussion. M. Kuhner, M. Stephens, S. Schaeffer and G. McVean generously assisted with computer programs. We thank W. G. Hill for editorial assistance and B. Jennings, A. Shedlock, H. Walsh, C. Welch and two anonymous reviewers provided very helpful comments on the manuscript. This research was supported by grants from the US National Science Foundation.

Appendix. Inferred haplotypes from Red-Winged Blackbird for the loci used in this study

The set of individuals sampled for the β-actin locus only partially overlaps those for the other loci. Localities are as follows: FL. Florida; NIC, Nicaragua; CO, Colorado; OR, Oregon; WA, Washington state; SE, south-east USA (Florida/Kentucky); NY, New York. Subscript numbers indicate the level of significance of inferred phase at that site from a typical run of PHASE when analysed simultaneously, as follows: 5, $<0{\cdot}61$; 6, 0·61–0·70; 7, 0·71–0·80; 8, 0·81–0·90; 9, 0·91–0·99. No subscript indicates certainty of phase or homozygosity.

(*A*)

Individual no.	Locality	Haplotype	40 kb locus sites														*Agph-DAB3* exon 2		*Agph-DAB1* promoter		
			76	94	135	155	205	206	225	263	301	321	330	337	397	478	15875	15904	40868	40900	40911
7	FL	1	G	G	T	A	G	A	G	T	C	G	C	G	A	C	A	T	A	G	T
		5	G	G	T	A	G	A	G	C	C	G	C	C	A	C	G	T	A	G	T
8	FL	8	G	G	T_6	A_6	G	A	G	C_9	T	G_9	C_9	C	A	C_6	G	T	G_5	G	T
		11	G	G	C_6	G_6	G	A	G	T_9	T	A_9	G_9	C	A	T_6	G	T	A_5	G	T
1004	NIC	4	G	G_7	T	A_7	G_8	A_6	G_8	T	T	A	G	C	A_6	T	G	T	G_6	G	T
		15	G	T_7	T	G_7	A_8	G_6	A_8	T	T	A	G	C	G_6	T	G	T	A_6	G	T
2005	CO	3	G	G	T	A	G	A	G	T_9	T_8	A_9	G_9	C	A	T_7	G	T	A_5	G_5	T_5
		6	G	G	T	A	G	A	G	C_9	C_8	G_9	C_9	C	A	C	G	T	G_5	A_5	G_5
2006	CO	2	G	G_6	T	A_6	G_7	A_8	G_7	T	C	A	G	C	A_6	T_7	G	T_6	A	G	T
		14	G	T_6	T	G_6	A_7	G_8	A_7	T	C	A	G	C	G_6	T	G	G_6	A	G	T
2020	CO	10	G	G_7	T	A_7	A_8	A	A_8	T	T	A	G	C	A_6	C_6	G	G	A	G	T
		13	G	T_7	T	G_7	G_8	A	G_8	T	T	A	G	C	G_6	T_6	G	G	A	G	T
2021	CO	5	G	G_6	T	A_7	G	A	G	C_9	C_8	G_9	C_9	C	A_9	C_7	G	T	A_5	G_5	T
		12	G	T_6	T	G_7	G	A	G	T_9	T_8	A_9	G_9	C	G_9	T_7	G	T	G_5	A_5	T
2022	CO	7	G_7	G	T_7	A_7	G	A	G	C_8	C	G_8	C_8	C	A	T	G	T	G_6	G	T
		16	A_7	G	C_7	G_7	G	A	G	T_8	C	A_8	G_8	C	A	T	G	T	A_6	G	T
243	OR	9	G_8	G	T_8	A_8	A_7	A	A_7	T	T	A	G	C	A_6	C_6	G	T	A	G	T
		17	A_8	G	C_8	G_8	G_7	A	G_7	T	T	A	G	C	G_6	T_6	G	T	A	G	T

(*B*) *β-fibrinogen intron 7/rhodopsin intron 2*

Individual no.	Haplotype	*BFIB7* sites																		*RDP2*	
		49	51	71	108	113	152	192	204	210	248	275	276	332	344	353	361	387	399	66	175
7	1	A	C	G	G	C	A	A	T	A	T	C	T	A	A	A	T	A	A	G	G
	1	A	C	G	G	C	A	A	T	A	T	C	T	A	A	A	T	A	A	G	G
8	6	A	C	G_7	G	C	A	A_7	T	G_8	G	C	C_8	A	G	T_5	T	T_5	A	G	G
	10	A	C	A_7	G	C	A	T_7	C	A_8	T	T	T_8	A	G	A_5	T	A_5	A	G	G
1004	7	A	C	G	G	C	T	A	T	A	G	C	T_5	A	G	A	T	A	A	G	G
	8	A	C	G	G	C	T	A	T	A	G	C	C_5	A	G	T	T	T	A	G	G
2005	4	A	C	G	G	C	A	A	T	G	G	C	T	A	G	A	T	A	A	G	G
	5	A	C	G	G	C	A	A	T	G	G	C	C	A	G	A	T	A	A	G	G
2006	2	A	C_7	G	G	C_7	A	A	T	A	T	C	T	A	A	A	T_7	A	G	T	G
	12	A	T_7	G	G	T_7	A	A	T	A	T	C	T	A	A	A	C_7	A	G	G	G
2020	9	A_5	C_7	A_7	G_7	C	A	A	C	A_8	T	T	T_8	A_5	A	A	T	A	A_7	G	G
	13	G_5	T_7	G_7	G_7	C	A	A	T	A_8	T	C	C_8	A_5	A	A	T	A	G_7	G	G
2021	5	A	C	G_7	G	C	A	A_7	T	G_8	G	C	C_7	A	G	A	T	A	A	G	A
	10	A	C	A_7	G	C	A	T_7	C	A_8	T	T	T_7	A	G	A	T	A	A	G	G
2022	2	A	C	G_7	G_6	C	A	A	T_8	A	T	C_8	T	A_6	A_8	A	T	A_7	G_7	G	G
	11	A	C	A_7	A_6	C	A	A	C_8	A	T	T_8	T	T_6	G_8	A	T	T_7	A_7	G	G
243	3	A	C	G_7	G	C	A	A	T	A	G_6	C	T	A	A	A	T	A	A	G	G
	9	A	C	A_7	G	C	A	A	C	A	T_6	T	T	A	A	A	T	A	A	G	G

(*C*) *β-actin intron 2*

			BACT2 sites							
Individual no.	Locality	Haplotype	37	55	104	119	176	215	272	304
1	WA	3	C_8	G	G	T_7	G	G	C	G_8
		7	A_8	G	G	C_7	G	G	C	C_8
2	WA	1	C	G	G	C	G	G	C	G
		3	C	G	G	T	G	G	C	G
3	WA	3	C_8	G	G	T_7	G	G	C	G_8
		7	A_8	G	G	C_7	G	G	C	C_8
4	WA	1	C	G	G	C	G	G	C	G
		1	C	G	G	C	G	G	C	G
5	WA	1	C_6	G	G	C	G	G	C	G
		7	A_6	G	G	C	G	G	C	C
6	WA	1	C	G	G	C	G	G	C	G
		1	C	G	G	C	G	G	C	G
7	SE	4	C_7	G	A_6	T_8	G	G	C	G
		6	A_7	G	G_6	C_8	G	G	C	G
8	SE	1	C	G	G	C	G	G	C	G
		1	C	G	G	C	G	G	C	G
9	SE	1	C	G	G	C	G	G	C	G
		6	A	G	G	C	G	G	C	G
10	NY	1	C_7	G	G	C	G	G_7	C_7	G
		9	A_7	G	G	C	A	A_7	C_7	G
11	NY	1	C	G	G	C	G	G	C	G
		6	A	G	G	C	G	G	C	G
12	NY	5	C_5	A	G	C	G	G	T	G
		8	A_5	G	G	C	G	G	T	G
13	SE	3	C_6	G	G	T	G	G	C	G
		6	A_6	G	G	C	G	G	C	G
14	SE	2	C	G	G	C	G	G	T	G
		8	A	G	G	C	G	G	T	G
15	SE	1	C	G	G	C	G	G	C	G
		6	A	G	G	C	G	G	C	G

References

Andolfatto, P. & Przeworski, M. (2000). A genome-wide departure from the standard neutral model in natural populations of Drosophila. *Genetics* **156**, 257–268.

Aquadro, C. F., Begun, D. J. & Kindahl, E. C. (1994). Selection, recombination and DNA polymorphism in *Drosophila*. In *Non-neutral Evolution: Theories and Molecular Data* (ed. B. Golding), pp. 46–56. New York: Chapman and Hall.

Axelsson, E., Smith, N. G., Sundstrom, H., Berlin, S. & Ellegren, V. (2004). Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and turkey. *Molecular Biology and Evolution* **21**, 1538–1547.

Ball, R. M., Jr, Freeman, S., James, F. C., Bermingham, E. & Avise, J. C. (1988). Phylogeographic population structure of red-winged blackbirds assessed by mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA* **85**, 1558–1562.

Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London, Series B* **355**, 1553–1562.

Begovich, A. B., McClure, G. R., Suraj, V. C., Helmuth, R. C., Fildes, N., Bugawan, T. L., Erlich, H. A. & Klitz, W. (1992). Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *Journal of Immunology* **148**, 249–258.

Begovich, A. B., Moonsamy, P. V., Mack, S. J., Barcellos, L. F., Steiner, L. L., Grams, S., Suraj Baker, V., Hollenbach, J., Trachtenberg, E., Louie, L., Zimmerman, P., Hill, A. V. S., Stoneking, M., Sasazuki, T., Konenkov, V. I., Sartakova, M. L., Titanji, V. P. K., Rickards, O. & Klitz, W. (2001). Genetic variability and linkage disequilibrium within the HLA-DP region: analysis of 15 different populations. *Tissue Antigens* **57**, 424–439.

Beletsky, L. (1996). *The Red-Winged Blackbird: The Biology of a Strongly Polygynous Songbird*. San Diego: Academic Press.

Bergström, T. P., Josefsson, A., Erlich, H. A. & Gyllensten, U. (1998). Recent origin of *HLA-DRB1* alleles and implications for human evolution. *Nature Genetics* **18**, 237–242.

Brumfield, R., Nickerson, D., Beerli, P. & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* **18**, 249–256.

Burt, D. W. (2002). Origin and evolution of avian microchromosomes. *Cytogenetics and Genome Research* **96**, 97–112.

Clark, A. G., Nielsen, R., Signorovitch, J., Matise, T. C., Glanowski, S., Heil, J., Winn-Deen, E. S., Holden, A. L. & Lai, E. (2003). Linkage disequilibrium and inference of ancestral recombination in 538 single-nucleotide polymorphism clusters across the human genome. *American Journal of Human Genetics* **73**, 285–300.

Congdon, B. C., Piatt, J. F., Martin, K. & Friesen, V. L. (2000). Mechanisms of population differentiation in marbled murrelets: historical versus contemporary processes. *Evolution* **54**, 974–986.

Crawford, D. C., Bhangale, T., Li N., Hellenthal, G., Rieder, M. J., Nickerson, D. A. & Stephens, M. (2004). Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nature Genetics* **36**, 700–706.

Dean, M., Carrington, M. & O'Brien, S. J. (2002). Balanced polymorphism selected by genetic versus infectious human disease. *Annual Review of Genomics and Human Genetics* **3**, 263–292.

Edwards, S. V. & Hedrick, P. W. (1998). Evolution and ecology of *MHC* molecules: from genomics to sexual selection. *Trends in Ecology and Evolution* **13**, 305–311.

Edwards, S., Gasper, J. & Stone, M. (1998). Genomics and polymorphism of *Agph-DAB1*, an *Mhc* class II B gene in Red-winged Blackbirds (*Agelaius phoeniceus*). *Molecular Biology and Evolution* **15**, 236–250.

Edwards, S. V., Gasper, J., Garrigan, D., Martindale, D. A. & Koop, B. F. (2000). A 39-kb sequence around a blackbird *Mhc* class II B gene: ghost of selection past and songbird genome architecture. *Molecular Biology and Evolution* **17**, 1384–1395.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* **78**, 737–756.

Freeman, S. & Zink, R. M. (1995). A phylogenetic study of the blackbirds based on variation in mitochondrial DNA restriction sites. *Systematic Biology* **44**, 409–420.

Garrigan, D. & Edwards, S. V. (1999). Polymorphism across an intron – exon boundary in an avian *Mhc* class II B gene. *Molecular Biology and Evolution* **16**, 1599–1606.

Garrigan, D. & Hedrick, P. W. (2003). Perspective. Detecting adaptive molecular polymorphism: lessons from the *MHC*. *Evolution* **57**, 1707–1722.

Gasper, J., Shiina, T., Inoko, H. & Edwards, S. V. (2001). Songbird genomics: analysis of 45-kb upstream of a polymorphic *Mhc* class II gene in Red-winged Blackbird (*Agelaius phoeniceus*). *Genomics* **75**, 26–34.

Gaudieri, S., Kulski, J. K., Dawkins, R. L. & Gojobori, T. (1999). Extensive nucleotide variability within a 370 kb sequence from the central region of the major histocompatibility complex. *Gene* **238**, 157–161.

Gaudieri, S., Dawkins, R. L., Habara, K., Kulski, J. K. & Gojobori, T. (2000). SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Research* **10**, 1579–1586.

Grafen, A. (1990). Do animals really recognize kin? *Animal Behaviour* **39**, 42–54.

Grimsley, C., Mather, K. A. & Ober, C. (1998). HLA-H: a pseudogene with increased variation due to balancing selection at neighboring loci. *Molecular Biology and Evolution* **15**, 1581–1588.

Gu, X. & Nei, M. (1999). Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian *MHC* genes. *Molecular Biology and Evolution* **16**, 147–156.

Gyllensten, U. B., Lashkari, D. & Erlich, H. A. (1990). Allelic diversification at the class II DQB locus of the mammalian major histocompatibility complex. *Proceedings of the National Academy of Sciences of the USA* **87**, 1835–1839.

Hala, K., Chaussé, A.-M., Bourlet, Y., Lassila, O., Hasler, V. & Auffrey, C. (1988). Attempt to detect recombination between B-F and B-L genes within the chicken B complex by serological typing, *in vitro* MLR and RFLP analyses. *Immunogenetics* **28**, 433.

Hare, M. (2001). Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* **16**, 700–706.

Hedrick, P. W. & Thomson, G. (1983). Evidence for balancing selection at HLA. *Genetics* **104**, 449–456.

Hedrick, P. W., Klitz, W., Robinson, W., Kuhner, M. K. & Thomson, G. (1991). Population genetics of HLA. In *Evolution at the Molecular Level* (ed. R. Selander, A. Clark & T. Whittam), pp. 248–271. Sunderland, MA, Sinauer.

Hellmann, I., Ebersberger, I., Ptak, S. E., Paabo, S. & Przeworski, M. (2003). A neutral explanation for the correlation of diversity with recombination rates in humans. *American Journal of Human Genetics* **72**, 1527–1535.

Hey, J. & Kliman, R. M. (2002). Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**, 595–608.

Hill, A. V. S. (1998). The immunogenetics of human infectious diseases. *Annual Review of Immunology* **16**, 593–617.

Hogstrand, K. & Bohme, J. (1999*a*). Gene conversion can create new *MHC* alleles. *Immunological Reviews* **167**, 305–317.

Hogstrand, K. & Bohme, J. (1999*b*). Gene conversion of major histocompatibility complex genes is associated with CpG-rich regions. *Immunogenetics* **49**, 446–455.

Hollenbach, J. A., Thomson, G., Cao, K., Fernandez Vina, M., Erlich, H. A., Bugawan, T. L., Winkler, C., Winter, M. & Klitz, J. (2001). HLA diversity, differentiation, and haplotype evolution in mesoamerican natives. *Human Immunology* **62**, 378–390.

Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics* **159**, 1805–1817.

Hudson, R., Kreitman, M. & Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hunt, H. D., Pharr, G. T. & Bacon, L. D. (1993). Molecular analysis reveals *MHC* class I intra-locus recombination in the chicken. *Immunogenetics* **40**, 370–375.

Jacob, J. P., Milne, S., Beck, S. & Kaufman, J. (2000). The major and a minor class II beta-chain (B-LB) gene flank the Tapasin gene in the B-F/B-L region of the chicken major histocompatibility complex. *Immunogenetics* **51**, 138–147.

Jacob, S., McClintock, M. K., Zelano, B. & Ober, C. (2002). Paternally inherited HLA alleles are associated with women's choice of male odor. *Nature Genetics* **30**, 175–179.

Jeffreys, A. J., Kauppi, L. & Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics* **29**, 217–222.

Kaufman, J. (1995). A 'minimal essential *Mhc*' and an 'unrecognized' *Mhc*: two extremes in selection for polymorphism. *Immunological Reviews* **143**, 63–88.

Kaufman, J. (1999). Co-evolving genes in *MHC* haplotypes: the 'rule' for nonmammalian vertebrates? *Immunogenetics* **50**, 228–236.

Kaufman, J. & Salomonsen, J. (1997). The 'Minimal Essential *MHC*' revisited: both peptide-binding and cell surface expression level of *MHC* molecules are polymorphisms selected by pathogens in chickens. *Hereditas* **127**, 67–73.

Kaufman, J., Milne, S., Gobel, T. W., Walker, B. A., Jacob, J. P., Auffray, C., Zoorob, R. & Beck, S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**, 923–925.

Klein, J., Satta, Y., O'Huigin, C. & Takahata, N. (1993). The molecular descent of the major histocompatibility complex. *Annual Review of Immunology* **11**, 269–295.

Kliman, R. M., Andolfatto, P., Coyne, J. A., Depaulis, F., Kreitman, M., Berry, A. J., McCarter, J., Wakeley, J. & Hey, J. (2000). The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913–1931.

Kuhner, M. K., Yamato, J. & Felsenstein, J. (2000). Maximum likelihood estimation of recombination rates from population data. *Genetics* **156**, 1393–1401.

Lercher, M. J. & Hurst, L. D. (2002). Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in Genetics* **18**, 337–340.

Lewontin, R. C. (1995). The detection of linkage disequilibrium in molecular sequence data. *Genetics* **140**, 377–388.

Li, N. & Stephens, M. (2003). Modelling linkage disequilibrium, and identifying recombination hotspots using SNP data. *Genetics*, **165**, 2213–2233.

Machado, C. A., Kliman, R. M., Markert, J. A. & Hey, J. (2002). Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution* **19**, 472–488.

Manning, C. J., Wakeland, E. K. & Potts, W. K. (1992). Communal nesting patterns in mice implicate *MHC* genes in kin recognition. *Nature* **360**, 581–583.

Maynard-Smith, J. & Haigh, J. (1974). The hitchhiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

McQueen, H. A., Siriaco, G. & Bird, A. P. (1998). Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Research* **8**, 621–630.

McVean, G., Awadalla, P. & Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241.

McVean, G. A., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R. & Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584.

Meager, S. & Potts, W. K. (1997). A microsatellite-based *MHC* genotyping system for house mice (*Mus domesticus*). *Hereditas* **127**, 75–82.

Nachman, M. W. (2001). Single nucleotide polymorphisms and recombination rate in humans. *Trends in Genetics* **17**, 481–485.

Nachman, M. W. (2002). Variation in recombination rate across the genome: evidence and implications. *Current Opinion in Genetics and Development* **12**, 657–663.

Navarro, A. & Barton, N. H. (2002). The effects of multilocus balancing selection on neutral variability. *Genetics* **161**, 849–863.

Nei, M. & Hughes, A. L. (1992). Balanced polymorphism and evolution by the birth-and-death process in the *MHC* loci. In *Proceedings of the 11th Histocompatibility Workshop and Conference* (ed. K. Tsuji, M. Aizawa & T. Sasazuki), pp. 27–38. Oxford: Oxford University Press.

Nickerson, D. A., Tobe, V. O. & Taylor, S. L. (1997). PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research* **25**, 2745–2751.

Nielsen, R. & Signorovitch, J. (2003). Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology* **63**, 245–255.

Ohta, T. (1999). Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunological Reviews* **167**, 319–325.

Ohta, T. (2000). An attempt to measure the patchwork pattern observed among alleles at major histocompatibility complex loci. *Journal of Molecular Evolution* **51**, 21–25.

Payseur, B. A. & Nachman, M. W. (2002*a*). Gene density and human nucleotide polymorphism. *Molecular Biology and Evolution* **19**, 336–340.

Payseur, B. A. & Nachman, M. W. (2002*b*). Natural selection at linked sites in humans. *Gene* **300**, 31–42.

Posada, D. (2002). Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Molecular Biology and Evolution* **19**, 708–717.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences of the USA* **98**, 13757–13762.

Posada, D., Crandall, K. A. & Holmes, E. C. (2002). Recombination in evolutionary genomics. *Annual Review of Genetics* **36**, 75–97.

Pritchard, J. K. & Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* **69**, 1–14.

Prychitko, T. M. & Moore, W. S. (1997). The utility of DNA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). *Molecular Phylogenetics and Evolution* **8**, 193–204.

Ptak, S. E., Roeder, A. D., Stephens, M., Gilad, Y., Pääbo, S. & Przeworski, M. (2004). Absence of the TAP2 human recombination hotspot in chimpanzees. *PLoS Biology* **2**, 849–855.

Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S. & Altshuler, D. (2002). Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genetics* **32**, 135–142.

Rice, W. R. (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Richman, A. D., Herrera, L. G. & Nash, D. (2003*a*). Evolution of *MHC* class II E beta diversity within the genus *Peromyscus*. *Genetics* **164**, 289–297.

Richman, A. D., Herrera, L. G., Nash, D. & Schierup, M. H. (2003*b*). Relative roles of mutation and recombination in generating allelic polymorphism at an *MHC* class II locus in *Peromyscus maniculatus*. *Genetical Research* **82**, 89–99.

Rieder, M. J., Taylor, S. L., Tobe, V. O. & Nickerson, D. A. (1998). Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Research* **26**, 967–973.

Rokas, A., Ladoukakis, E. & Zouros, E. (2003). Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution* **18**, 411–417.

Rozas, J. & Rozas, R. (1997). DNAsp Version 2.0: a novel software package for extensive molecular population genetics analysis. *Computer Applications in the Biosciences* **13**, 307–311.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Satta, Y., O'Huigin, C., Takahata, N. & Klein, J. (1994). Intensity of natural selection at the major histocompatibility complex loci. *Proceedings of the National Academy of Sciences of the USA* **91**, 7184–7188.

Satta, Y., Kupfermann, H., Li, Y. J. & Takahata, N. (1999). Molecular clock and recombination in primate *Mhc* genes. *Immunological Reviews* **167**, 367–379.

Schaeffer, S. W., Walthour, C. S., Toleno, D. M., Olek, A. T. & Miller, E. L. (2001). Protein variation in Adh and Adh-related in *Drosophila pseudoobscura*. Linkage disequilibrium between single nucleotide polymorphisms and protein alleles. *Genetics* **159**, 673–687.

She, J. X., Boehme, S., Wang, T. W., Bonhomme, F. & Wakeland, E. K. (1990). The generation of *MHC* class II gene polymorphism in the genus *Mus*. *Biological Journal of the Linnaean Society* **41**, 141–161.

Slatkin, M. (2000). Balancing selection at closely linked, overdominant loci in a finite population. *Genetics* **154**, 1367–1378.

Slatkin, M. & Muirhead, C. A. (2000). A method for estimating the intensity of overdominant selection from the distribution of allele frequencies. *Genetics* **156**, 2119–2126.

Slatkin, M. & Wiehe, T. (1998). Genetic hitch-hiking in a subdivided population. *Genetical Research* **71**, 155–160.

Smith, J. & Burt, D. W. (1998). Parameters of the chicken genome (*Gallus gallus*). *Animal Genetics* **29**, 290–294.

Stephens, M. & Donnelly, P. (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**, 1162–1169.

Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., Jiang, R., Messer, C. J., Chew, A., Han, J. H., Duan, J., Carr, J. L., Lee, M. S., Koshy, B., Kumar, A. M., Zhang, G., Newell, W. R., Windemuth, A., Xu, C., Kalbfleisch, T. S., Shaner, S. L., Arnold, K., Schulz, V., Drysdale, C. M., Nandabalan, K., Judson, R. S., Ruano, G. & Vovis, G. F. (2001*a*). Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**, 489–493.

Stephens, M., Smith, N. J. & Donnelly, P. (2001*b*). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978–989.

Stumpf, M. P. & McVean, G. A. (2003). Estimating recombination rates from population-genetic data. *Nature Reviews Genetics* **4**, 959–968.

Swofford, D. L. (1999). *PAUP* - Phylogenetic Analysis Using Parsimony*, version 4.0b. Sinauer Associates, Sunderland, MA.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–589.

Takahata, N. & Satta, Y. (1998). Footprints of intragenic recombination at *HLA* loci and human evolution. *Immunogenetics* **47**, 430–441.

Thomson, G. (1977). The effect of a selected locus on linked neutral loci. *Genetics* **85**, 753–788.

Wakeley, J. & Hey, J. (1997). Estimating ancestral population parameters. *Genetics* **145**, 847–855.

Wakeley, J. & Lessard, S. (2003). Theory of the effects of population structure and sampling on patterns of linkage disequilibrium applied to genomic data from humans. *Genetics* **164**, 1043–1053.

Wakeley, J., Nielsen, R., Liu-Cordero, S. N. & Ardlie, K. (2001). The discovery of single-nucleotide polymorphisms – and inferences about human demographic history. *American Journal of Human Genetics* **69**, 1332–1347.

Walsh, E. C., Mather, K. A., Schaffner, S. F., Farwell, L., Daly, M. J., Patterson, N., Cullen, M., Carrington, M., Bugawan, T. L., Erlich, H., Campbell, J., Barrett, J., Miller, K., Thomson, G., Lander, E. S. & Rioux, J. D. (2003). An integrated haplotype map of the human major histocompatibility complex. *American Journal of Human Genetics* **73**, 580–590.

Waltari, E. & Edwards, S. V. (2002). Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *American Naturalist* **160**, 539–552.

Wang, Z., Hill, G. E., Baker, A. J. & Edwards, S. V. (2003). Reconciling actual and inferred population histories in the House Finch (*Carpodacus mexicanus*) by AFLP analysis. *Evolution* **57**, 2852–2864.

Yeager, M. & Hughes, A. L. (1999). Evolution of the mammalian *MHC*: natural selection, recombination, and convergent evolution. *Immunological Reviews* **167**, 45–58.

Zelano, B. & Edwards, S. V. (2002) An Mhc component to kin recognition and mate choice in birds: predictions, progress, and prospects. *American Naturalist* **160**, S225–S237.